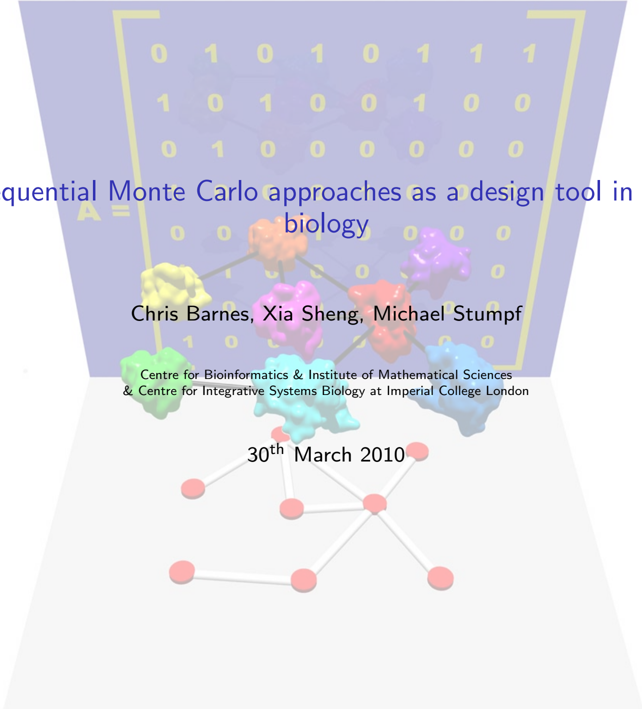


Using sequential Monte Carlo approaches as a design tool in synthetic biology

Chris Barnes, Xia Sheng, Michael Stumpf

Centre for Bioinformatics & Institute of Mathematical Sciences
& Centre for Integrative Systems Biology at Imperial College London

30th March 2010

Introduction

Synthetic genetic circuits

- Understand how organisms function

Introduction

Synthetic genetic circuits

- Understand how organisms function
- Produce drugs more effectively eg anti malarials
- Metabolize toxic chemicals
- Modify bacteria to hunt and kill tumors
- Stem cell production

Introduction

Synthetic genetic circuits

- Understand how organisms function
- Produce drugs more effectively eg anti malarials
- Metabolize toxic chemicals
- Modify bacteria to hunt and kill tumors
- Stem cell production

Challenges

- Tuning of co-operativity, repression strengths, decay rates
- Genetic components relatively unreliable with large variations in parameter values
- Interactions and cross-talk at the systems level
- Asynchronous timing

Introduction

Synthetic genetic circuits

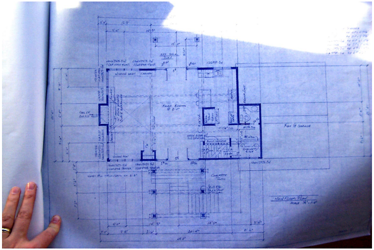
- Understand how organisms function
- Produce drugs more effectively eg anti malarials
- Metabolize toxic chemicals
- Modify bacteria to hunt and kill tumors
- Stem cell production

Challenges

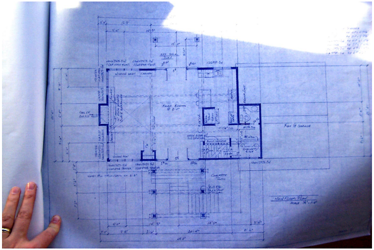
- Tuning of co-operativity, repression strengths, decay rates
- Genetic components relatively unreliable with large variations in parameter values
- Interactions and cross-talk at the systems level
- Asynchronous timing

How do we design genetic circuits to perform these functions?

Synthetic biology vs Systems biology



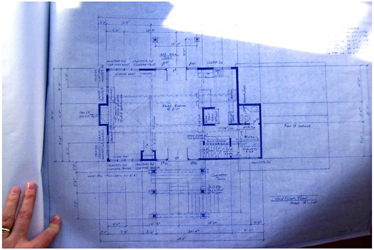
Synthetic biology vs Systems biology



System design as an inference problem

Which configuration of components will give an output O given an input I?
Which model best describes the observed data O given conditions I?

Synthetic biology vs Systems biology



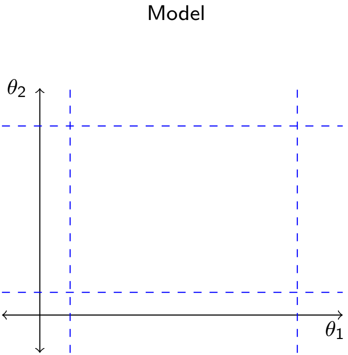
System design as an inference problem

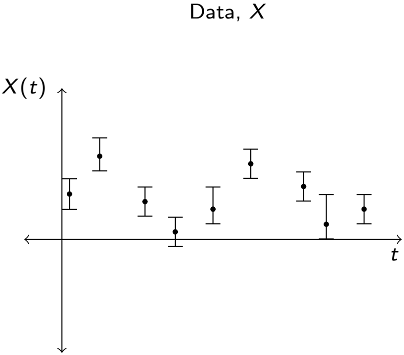
Which configuration of components will give an output O given an input I?
Which model best describes the observed data O given conditions I?

Advantages of using Bayesian framework for system design

- Credible limits on parameter values vs optimum values
- Model selection through Bayes factors
- Sensitivity from posterior distribution
- Incorporate prior biological knowledge into system design

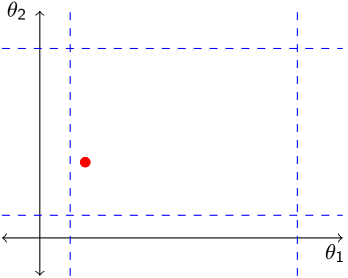
Approximate Bayesian Computation



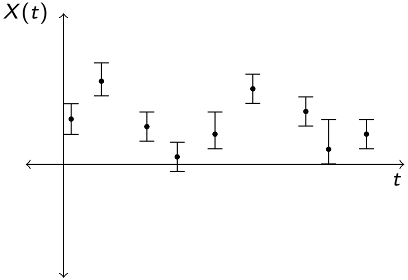


Approximate Bayesian Computation

Model

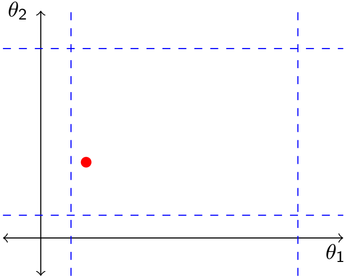


Data, X

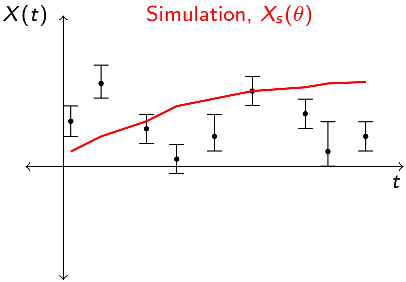


Approximate Bayesian Computation

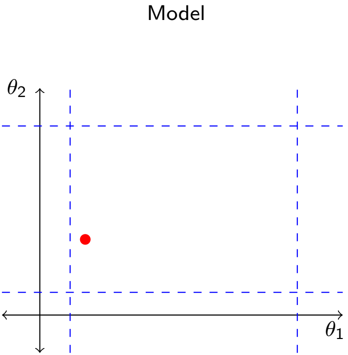
Model

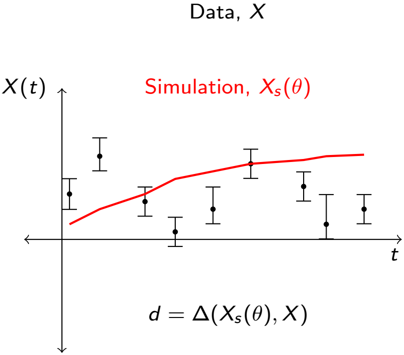


Data, X



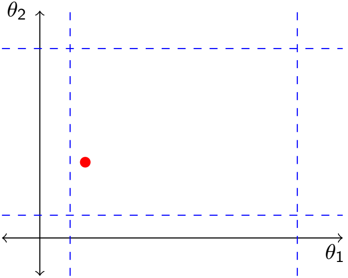
Approximate Bayesian Computation



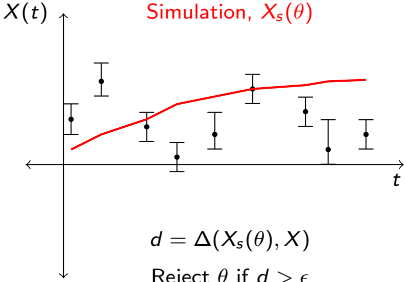


Approximate Bayesian Computation

Model



Data, X

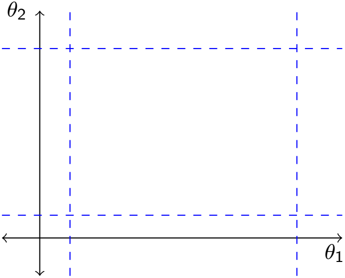


$d = \Delta(X_s(\theta), X)$

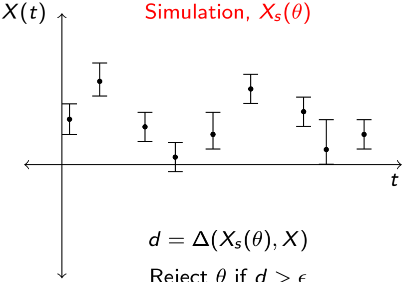
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

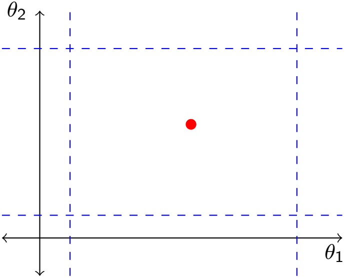


$$d = \Delta(X_s(\theta), X)$$

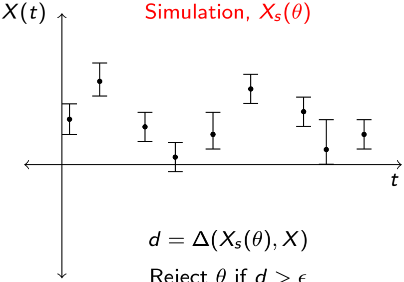
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



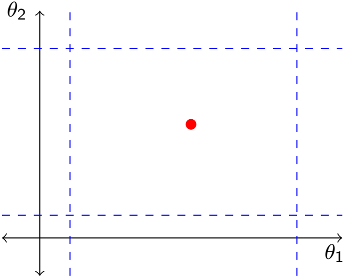
Data, X



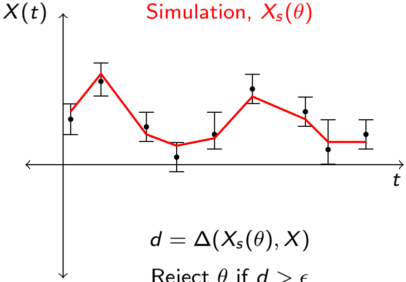
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

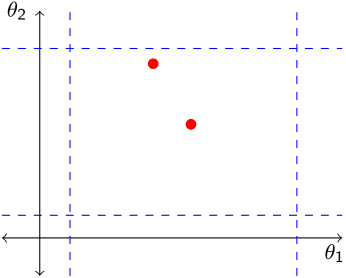


$$d = \Delta(X_s(\theta), X)$$

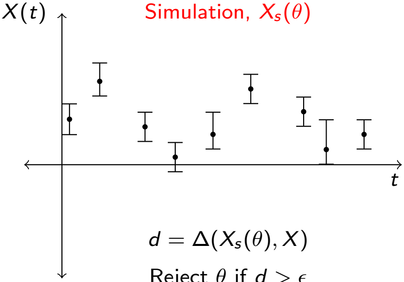
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

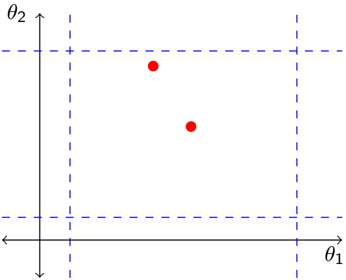


$$d = \Delta(X_s(\theta), X)$$

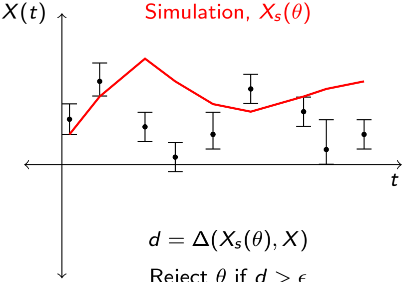
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

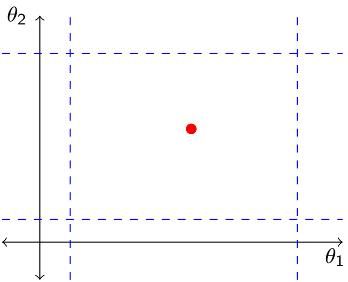


$$d = \Delta(X_s(\theta), X)$$

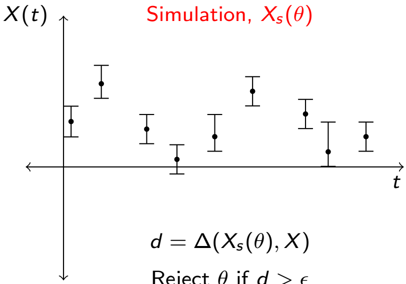
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

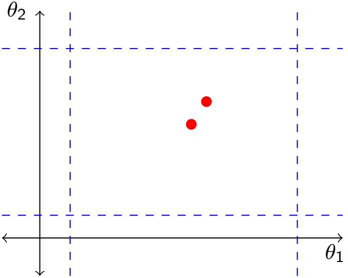


$d = \Delta(X_s(\theta), X)$

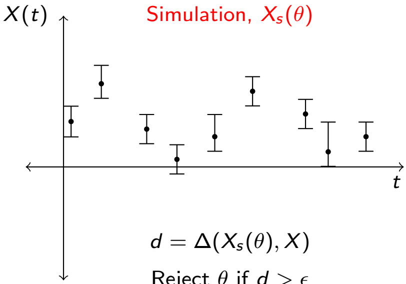
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

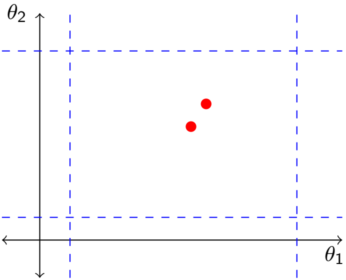


$$d = \Delta(X_s(\theta), X)$$

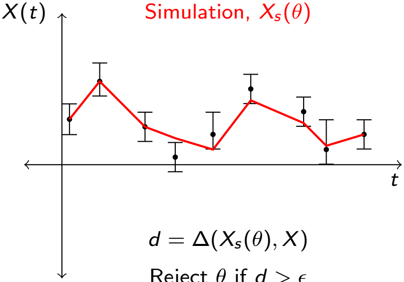
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

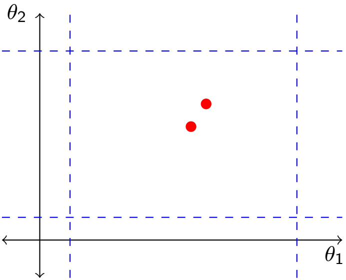


$d = \Delta(X_s(\theta), X)$

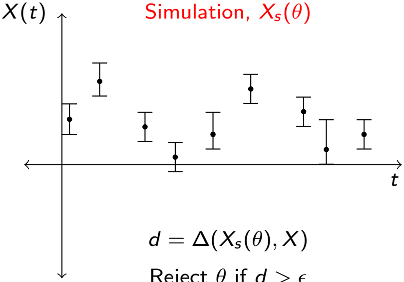
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



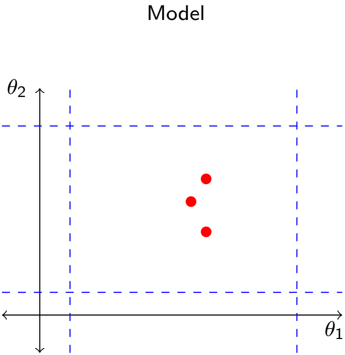
Data, X

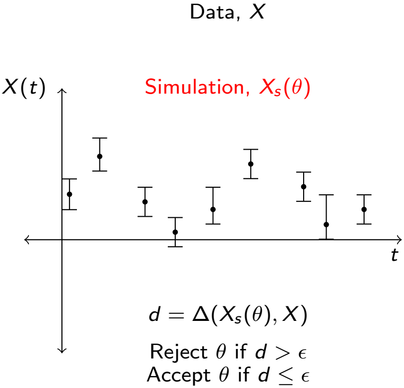


$d = \Delta(X_s(\theta), X)$

Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

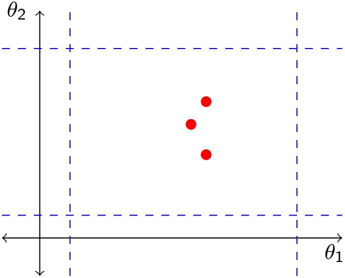
Approximate Bayesian Computation



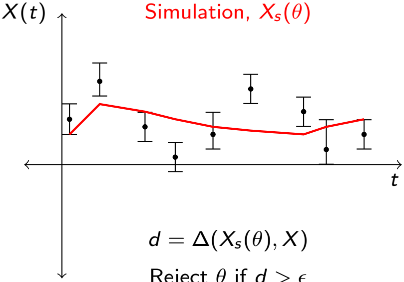


Approximate Bayesian Computation

Model



Data, X

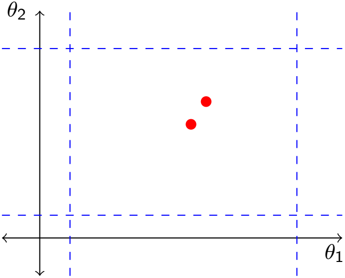


$$d = \Delta(X_s(\theta), X)$$

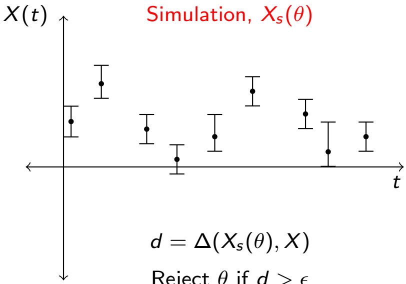
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

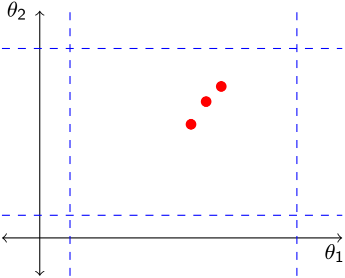


$d = \Delta(X_s(\theta), X)$

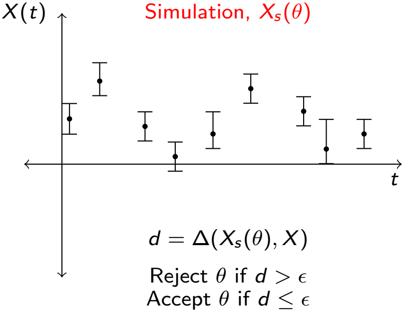
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model

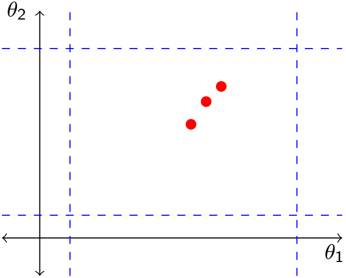


Data, X

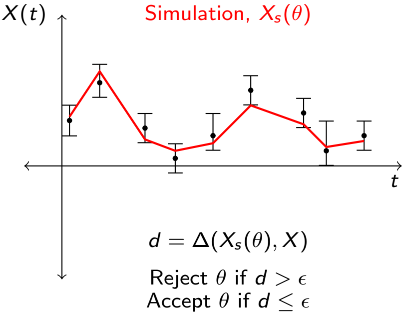


Approximate Bayesian Computation

Model

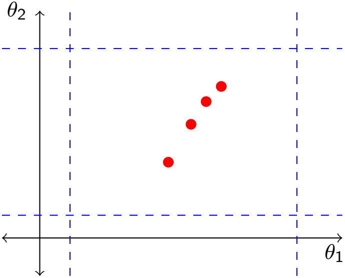


Data, X

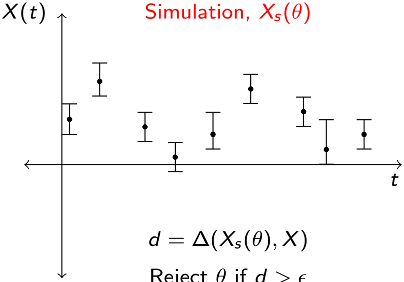


Approximate Bayesian Computation

Model



Data, X

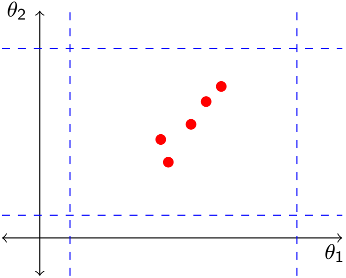


$$d = \Delta(X_s(\theta), X)$$

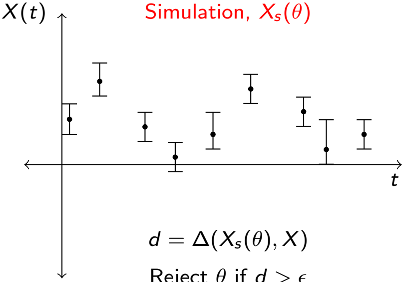
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



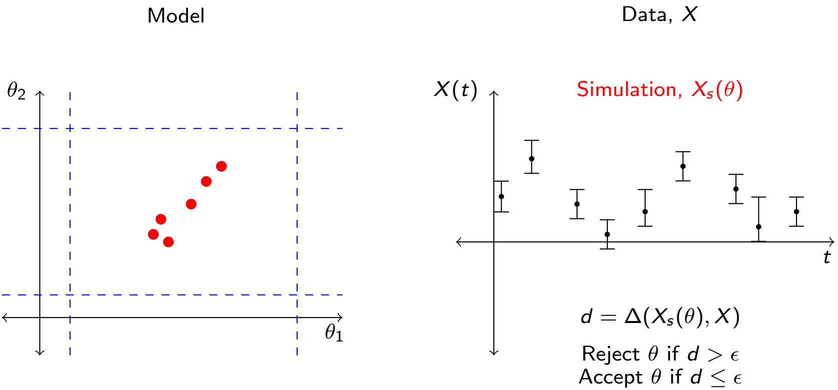
Data, X



$$d = \Delta(X_s(\theta), X)$$

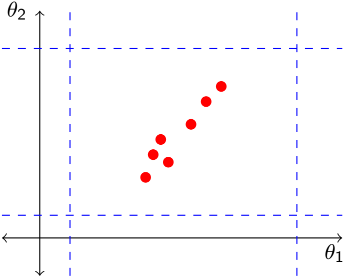
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

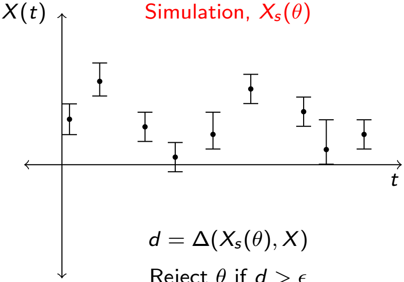


Approximate Bayesian Computation

Model



Data, X

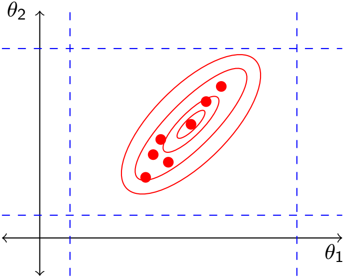


$$d = \Delta(X_s(\theta), X)$$

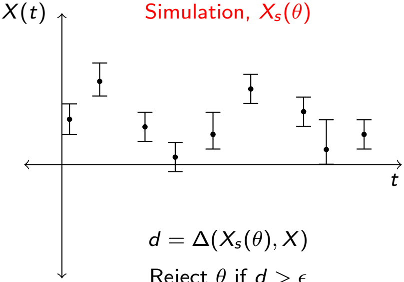
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X



$$d = \Delta(X_s(\theta), X)$$

Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation (ABC)

Bayesian Inference

Posterior	$\propto$	Likelihood	$\times$	prior
$p(\theta\|X)$	$\propto$	$p(X\|\theta)$		$p(\theta)$

Approximate inference methods

Sample from approximate posterior:

$$p(\theta|\Delta(X_s(\theta), X) \leq \epsilon).$$

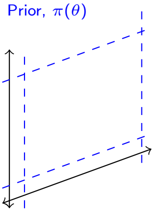
where $\Delta(X_s, X)$ is distance between simulation and data.
It can be shown, as $\epsilon \longrightarrow 0$

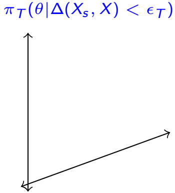
$$p(\theta|\Delta(X_s(\theta), X) \leq \epsilon) \longrightarrow p(\theta|X)$$

ABC flavours

- ABC rejection Pritchard *et al.* Mol. Biol. Evol. (1999)
- ABC MCMC Marjoram *et al.* PNAS (2003)
- ABC SMC Toni and Stumpf, Bioinformatics (2010)

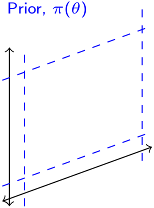
ABC SMC





ABC SMC

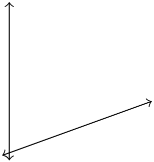
Prior, $\pi(\theta)$



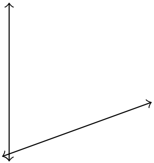
Define set of intermediate distributions, π_t, $t = 1, \ldots., T$

$$\epsilon_1 > \epsilon_2 > \ldots\ldots > \epsilon_T$$

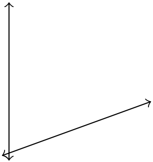
$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$



$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$

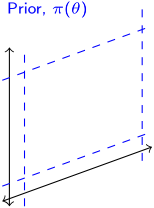


$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$



ABC SMC

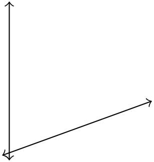
Prior, $\pi(\theta)$



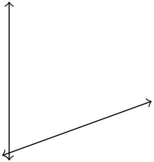
Define set of intermediate distributions, π_t, $t = 1, ..., T$

$$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$$

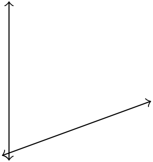
$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$



$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$



$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$



Sequential importance sampling:

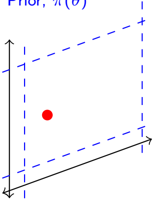
Sample from proposal, $\eta_t(\theta_t)$ and weight $w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$

$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$

$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

ABC SMC

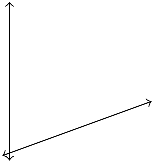
Prior, $\pi(\theta)$



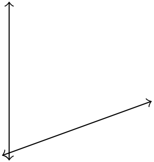
Define set of intermediate distributions, π_t, $t = 1, \ldots, T$

$$\epsilon_1 > \epsilon_2 > \ldots\ldots > \epsilon_T$$

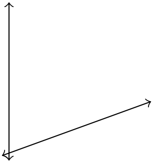
$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$



$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$



$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$



Sequential importance sampling:

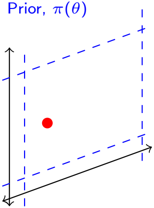
Sample from proposal, $\eta_t(\theta_t)$ and weight $w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$

$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$

$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

ABC SMC

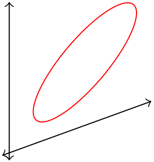
Prior, $\pi(\theta)$



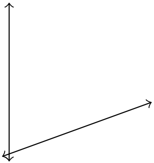
Define set of intermediate distributions, π_t, $t = 1, ..., T$

$$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$$

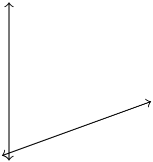
$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$



$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$



$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$



Sequential importance sampling:

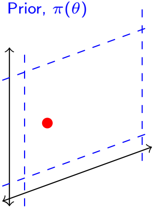
Sample from proposal, $\eta_t(\theta_t)$ and weight $w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$

$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$

$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

ABC SMC

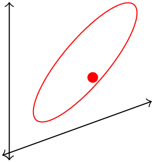
Prior, $\pi(\theta)$



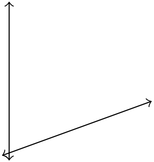
Define set of intermediate distributions, π_t, $t = 1, \ldots., T$

$$\epsilon_1 > \epsilon_2 > \ldots\ldots > \epsilon_T$$

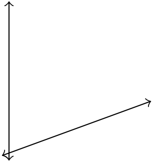
$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$



$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$



$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$



Sequential importance sampling:

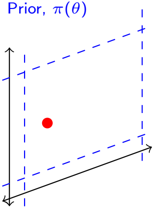
Sample from proposal, $\eta_t(\theta_t)$ and weight $w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$

$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1}) K_t(\theta_{t-1}, \theta_t) d\theta_{t-1}$

$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

ABC SMC

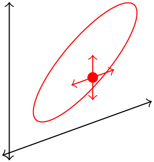
Prior, $\pi(\theta)$



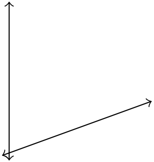
Define set of intermediate distributions, π_t, $t = 1, ..., T$

$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$

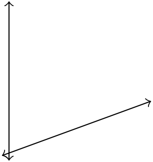
$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$



$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$



$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$



Sequential importance sampling:

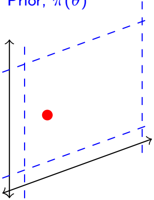
Sample from proposal, $\eta_t(\theta_t)$ and weight $w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$

$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$

$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

ABC SMC

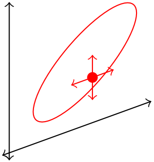
Prior, $\pi(\theta)$



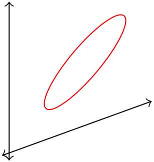
Define set of intermediate distributions, π_t, $t = 1, ..., T$

$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$

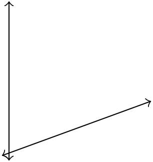
$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$



$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$



$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$



Sequential importance sampling:

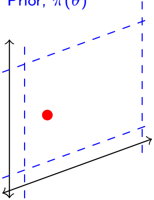
Sample from proposal, $\eta_t(\theta_t)$ and weight $w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$

$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$

$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

ABC SMC

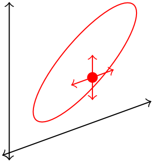
Prior, $\pi(\theta)$



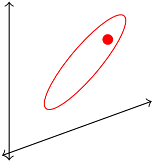
Define set of intermediate distributions, π_t, $t = 1, ..., T$

$$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$$

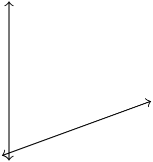
$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$



$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$



$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$



Sequential importance sampling:

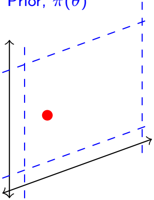
Sample from proposal, $\eta_t(\theta_t)$ and weight $w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$

$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$

$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

ABC SMC

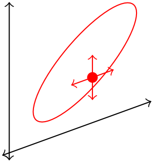
Prior, $\pi(\theta)$



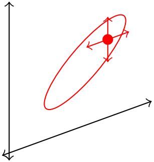
Define set of intermediate distributions, π_t, $t = 1, ..., T$

$$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$$

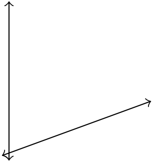
$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$



$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$



$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$



Sequential importance sampling:

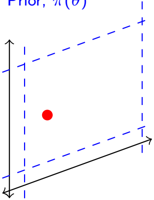
Sample from proposal, $\eta_t(\theta_t)$ and weight $w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$

$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1}) K_t(\theta_{t-1}, \theta_t) d\theta_{t-1}$

$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

ABC SMC

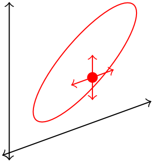
Prior, $\pi(\theta)$



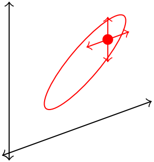
Define set of intermediate distributions, $\pi_t, t = 1, ..., T$

$$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$$

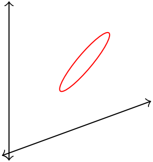
$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$



$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$



$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$



Sequential importance sampling:

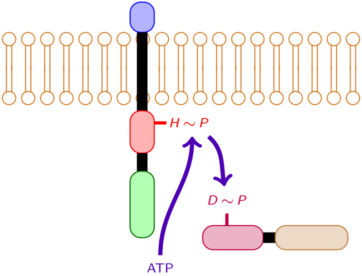
Sample from proposal, $\eta_t(\theta_t)$ and weight $w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$

$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1}) K_t(\theta_{t-1}, \theta_t) d\theta_{t-1}$

$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

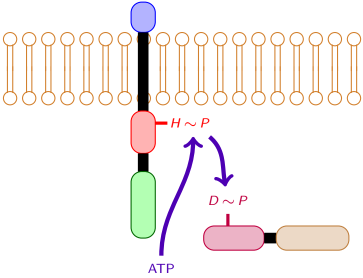
Bacterial two component systems (TCS)

- TCSs are abundant in bacteria, plants and fungi, but apparently absent from animals.
- They regulate response to environmental stimuli.



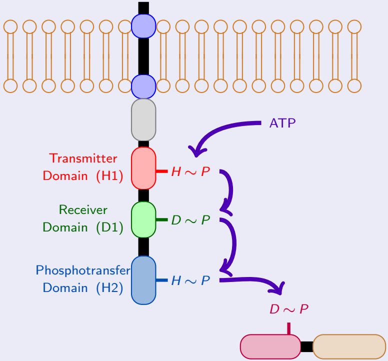
Bacterial two component systems (TCS)

- TCSs are abundant in bacteria, plants and fungi, but apparently absent from animals.
- They regulate response to environmental stimuli.



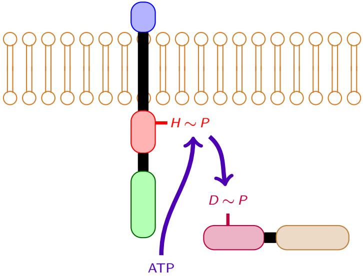
The ArcB-ArcA TCS

The ArcB-ArcA system in *Escherichia coli* uses a phospho-relay mechanism.



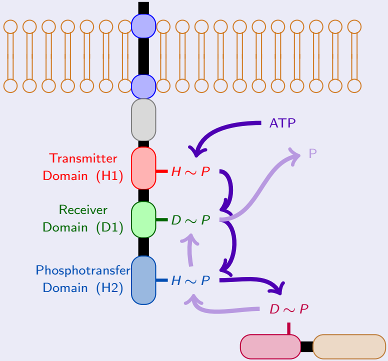
Bacterial two component systems (TCS)

- TCSs are abundant in bacteria, plants and fungi, but apparently absent from animals.
- They regulate response to environmental stimuli.



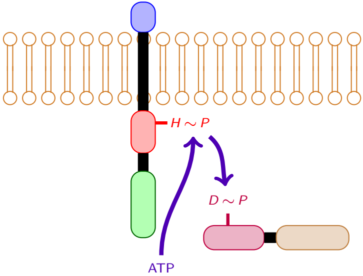
The ArcB-ArcA TCS

The ArcB-ArcA system in *Escherichia coli* uses a phospho-relay mechanism.



Bacterial two component systems (TCS)

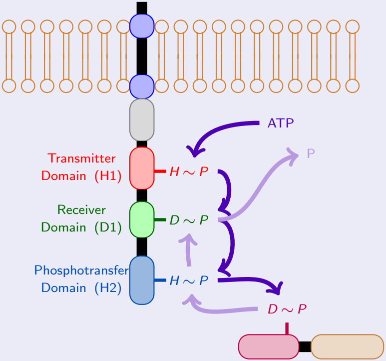
- TCSs are abundant in bacteria, plants and fungi, but apparently absent from animals.
- They regulate response to environmental stimuli.



Orthodox system

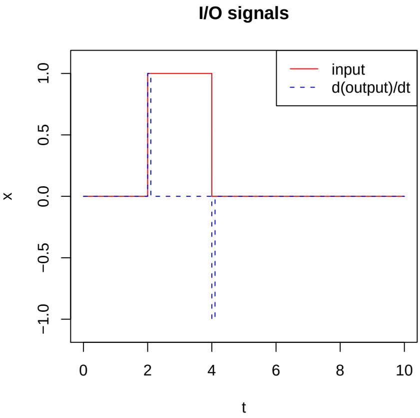
The ArcB-ArcA TCS

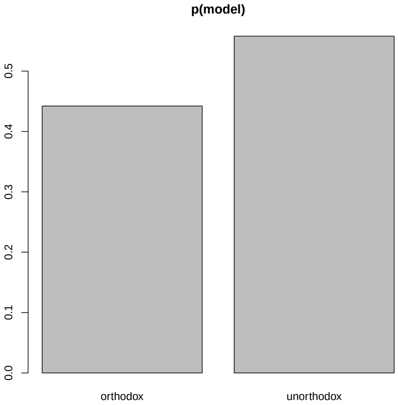
The ArcB-ArcA system in *Escherichia coli* uses a phospho-relay mechanism.



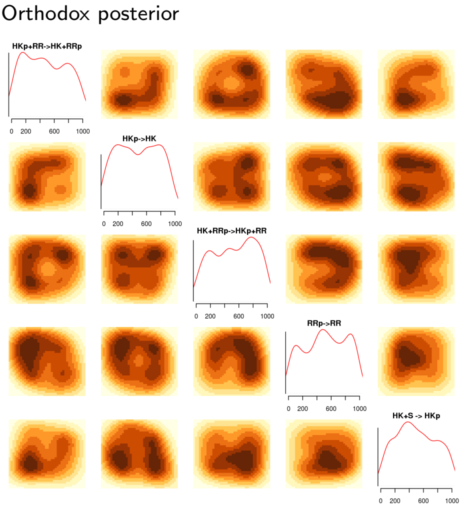
Unorthodox system

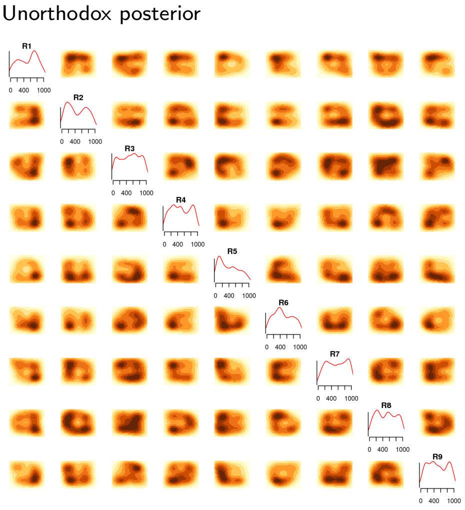
Example1 : Fast response





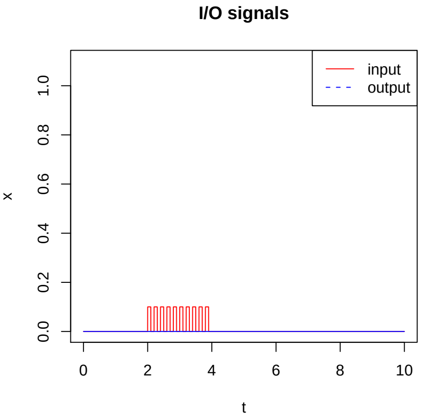
Example1 : Fast response

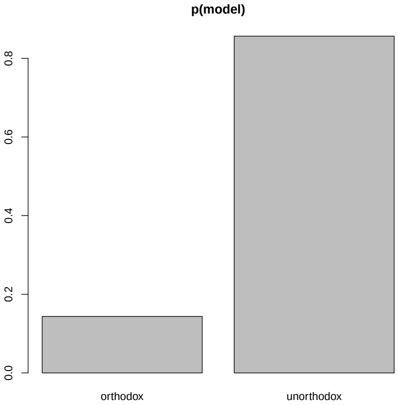




R1: $x \bullet \bullet \longrightarrow o \bullet \bullet$, R2: $ox\bullet \longrightarrow xo\bullet$, R3: $\bullet ox \longrightarrow \bullet xo$, R4: $\bullet \bullet o + RR \longrightarrow \bullet \bullet x + RRp$, R5: $\bullet xo \longrightarrow \bullet ox$ R6: $\bullet o \bullet \longrightarrow \bullet x \bullet$, R7: $\bullet \bullet x + RRp \longrightarrow \bullet \bullet o + RR$, R8: $o \bullet \bullet \longrightarrow x \bullet \bullet$, R9: $RRp \longrightarrow RR$

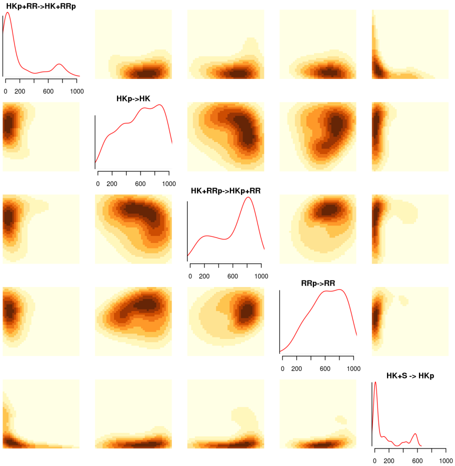
Example2 : Robust to noise



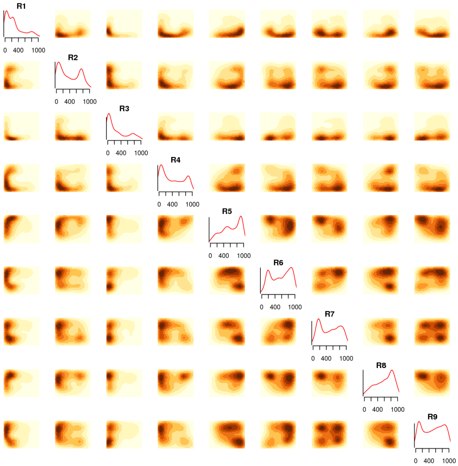


Example2 : Robust to noise

Orthodox posterior

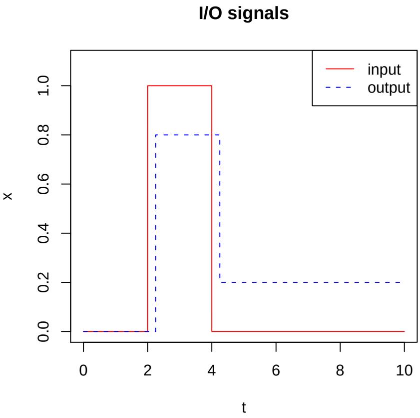


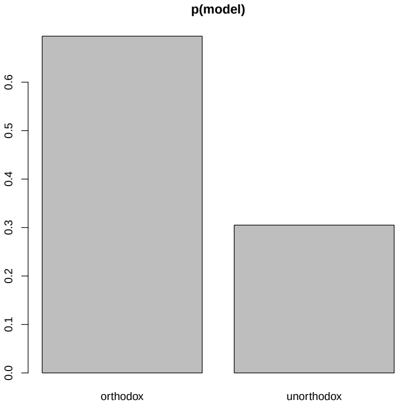
Unorthodox posterior



R1: $x \bullet \bullet \longrightarrow o \bullet \bullet$, R2: $ox\bullet \longrightarrow xo\bullet$, R3: $\bullet ox \longrightarrow \bullet xo$, R4: $\bullet \bullet o + RR \longrightarrow \bullet \bullet x + RRp$, R5: $\bullet xo \longrightarrow \bullet ox$ R6: $\bullet o \bullet \longrightarrow \bullet x \bullet$,
R7: $\bullet \bullet x + RRp \longrightarrow \bullet \bullet o + RR$, R8: $o \bullet \bullet \longrightarrow x \bullet \bullet$, R9: $RRp \longrightarrow RR$

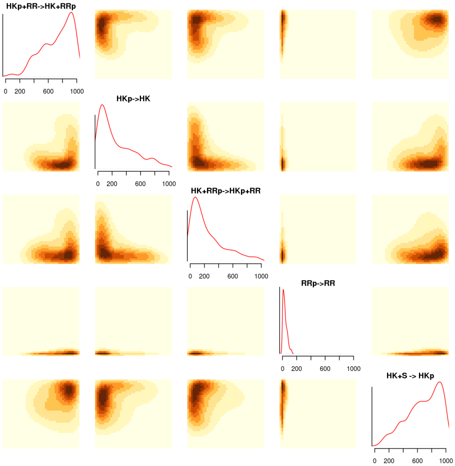
Example3 : Fast response, high maximum, low minimum



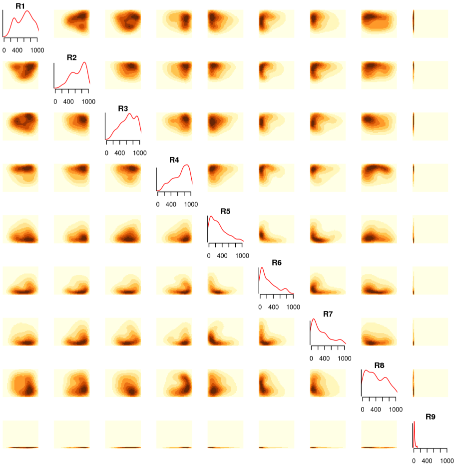


Example 3: posteriors

Orthodox posterior

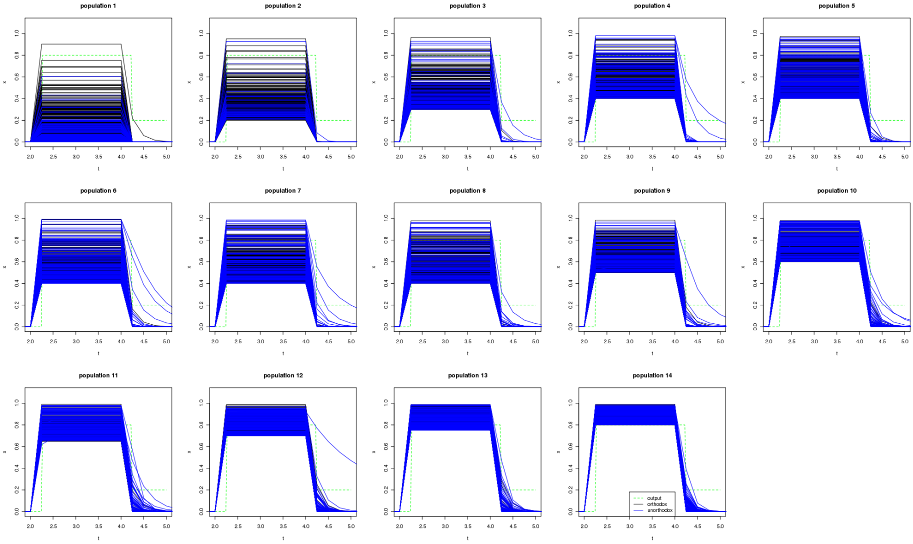


Unorthodox posterior



R1: $x \bullet \bullet \longrightarrow o \bullet \bullet$, R2: $ox\bullet \longrightarrow xo\bullet$, R3: $\bullet ox \longrightarrow \bullet xo$, R4: $\bullet \bullet o + RR \longrightarrow \bullet \bullet x + RRp$, R5: $\bullet xo \longrightarrow \bullet ox$ R6: $\bullet o \bullet \longrightarrow \bullet x \bullet$,
R7: $\bullet \bullet x + RRp \longrightarrow \bullet \bullet o + RR$, R8: $o \bullet \bullet \longrightarrow x \bullet \bullet$, R9: $RRp \longrightarrow RR$

Example 3: Trajectory evolution

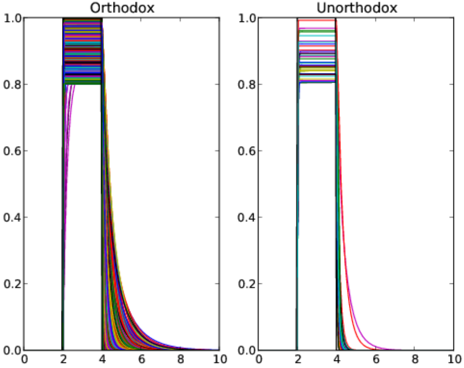


Example 3: Comparison to sensitivity analysis

Two approaches

- Examine region around optimum
- Sample uniformly from the prior

2060/10000	39/10000
0.981	0.019

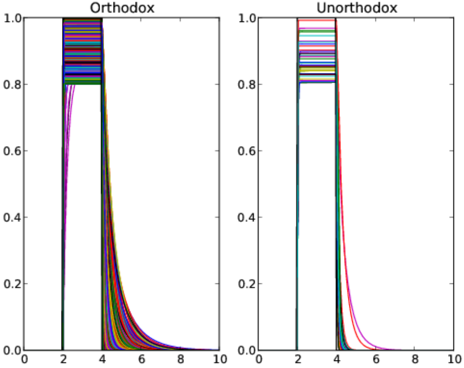


Example 3: Comparison to sensitivity analysis

Two approaches

- Examine region around optimum May miss alternative parameter combinations
- Sample uniformly from the prior

2060/10000	39/10000
0.981	0.019

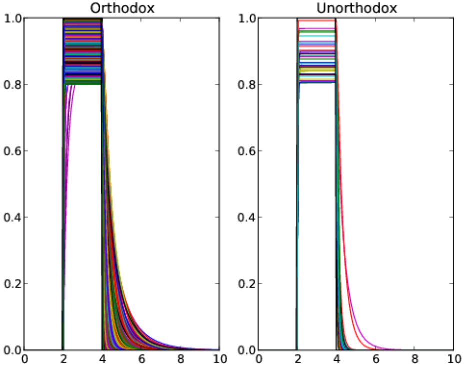


Example 3: Comparison to sensitivity analysis

Two approaches

- Examine region around optimum May miss alternative parameter combinations
- Sample uniformly from the prior May miss important regions

2060/10000	39/10000
0.981	0.019



Future work

BioBricks^{TM}

- Use selected BioBricks components to design genetic circuits and make predictions of behaviour under perturbations
- Build the circuit to test predictions

Genetic Design Automation (GDA)

- Genetic equivalent of Electronic Design Automation (EDA)
- Build complex genetic circuits and DNA sequence automatically from standard set of building blocks
- Engineering Genetic Circuits, Myers (2009)

Distant future

Using biological circuits to influence electronic circuit design

Acknowledgements

Michael Stumpf
Xia Sheng
Tina Toni
Maxime Huvet
Kamil Erguler
Justina Norkunaite
Juliane Liepe
Paul Kirk
Suhail Islam

bbsrc
biotechnology and biological sciences
research council

christopher.barnes@imperial.ac.uk

http://www3.imperial.ac.uk/theoreticalsystemsbiology
http://abc-sysbio.sourceforge.net/